U. Colorado at Colorado Springs

UCCS

Securics

Beyond Image Quality

Failure Analysis from Similarity Surface Techniques

Terrance Boult

El Pomar Professor of Communication and Computation

University of Colorado at Colorado Springs

And founder,

Securics Inc.

With past work by at Lehigh by R. Micheals, Weiliang Li, Yin Chen, Xiang Gao T. Riopkia,

At UCCS with Jay Potharaju

uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Recommendations

- Need to develop consistent measure of quality of "utility quality measures" that allow comparison.
  - We recommend FP ROC.
- Community should separate issues different "Qualities" and needs to work on at least 4 different "utility" qualities:
  - Capture, Enrollment, Match/Failure, Share
- Compared to finger matching, Data/features used by face algorithms has significantly greater variations, so cannot expect same "prediction" ability from image quality.
- Blind SNR estimates workable for image-quality. Can be improve by weighting "feature regions" and learning features for Eyes/Glasses/Pose.
- Can develop a general PRAT/FASST Toolkit for algorithm "match quality" from biometric algorithm specific data.

U. Colorado at Colorado Springs

UCCS

Securics

Blur

Atmosphere/Weather

Noise

How do sensor/world variations impact Face Recognition?

Compression

Gamma

Dynamic Range

Need controlled/designed experiments!

uccs.edu

U. Colorado at Colorado Springs

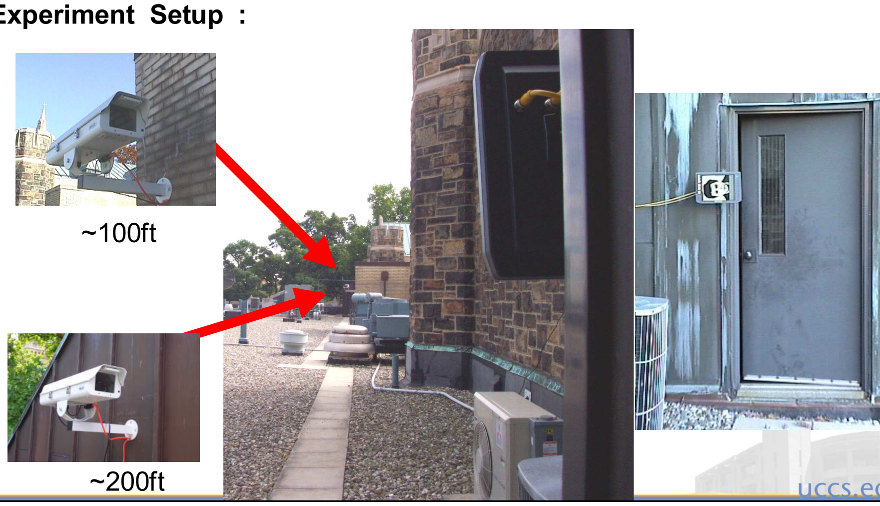
UCCS

Securics

Photo-head Data Acquisition

Sensor : FOV 0.5° and 0.25° imaging (equivalent to 1600mm and 3200mm focal lengths).

Experiment Setup :



~100ft

~200ft

uccs.edu

U. Colorado at Colorado Springs

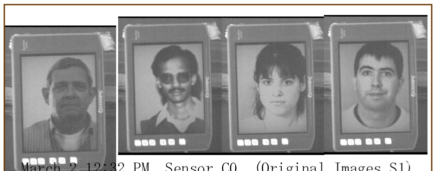
UCCS

Securics

Example Photoheads

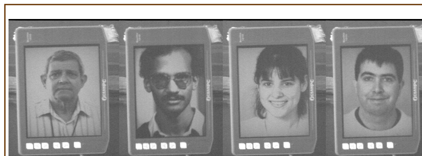
S1 Gallery

S2 Gallery



March 2 12:32 PM, Sensor C0, (Original Images S1)

(C0, S1) Probe Set



March 2 12:32 PM, Sensor C0, (Original Images S2)

(C0, S2) Probe Set

March 2 12:32 PM, Sensor C1, (Original Images S1)

(C1, S1) Probe Set

March 2 12:32 PM, Sensor C1, (Original Images S2)

(C1, S2) Probe Set

uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Example "photohead" data

100ft **200ft**

9:30am – 8:pm (4 samples per hour)

DARPA HID Conference, September 2002

U. Colorado at Colorado Springs

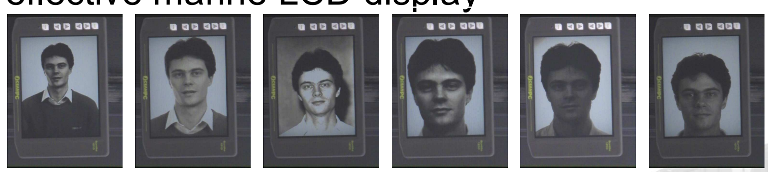
UCCS

Securics

Experiments

- Four datasets: JPEG, Outdoor, Blur, & Gamma
  - **JPEG**: Varying image quality from 100 to 0

  - **Outdoor**: Images collected from outdoor anti-reflective marine LCD display



DARPA HID — HBASE collection: Camera distance = 100 / 200ft

uccs.edu

U. Colorado at Colorado Springs

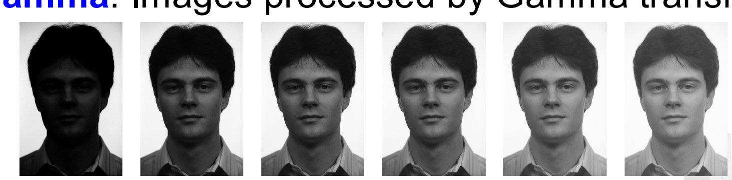
UCCS

Securics

Experiments

  - **Blur**: Blurred images by Gaussian kernel 7×7

  - **Gamma**: Images processed by Gamma transform



uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Facial Image Quality from blind SNR estimate

Statistical properties of edge image change with quality. Suppose pdf of edge intensity image, $||\nabla I||$ is $f_{||\nabla I||}(\cdot)$ has mean μ.

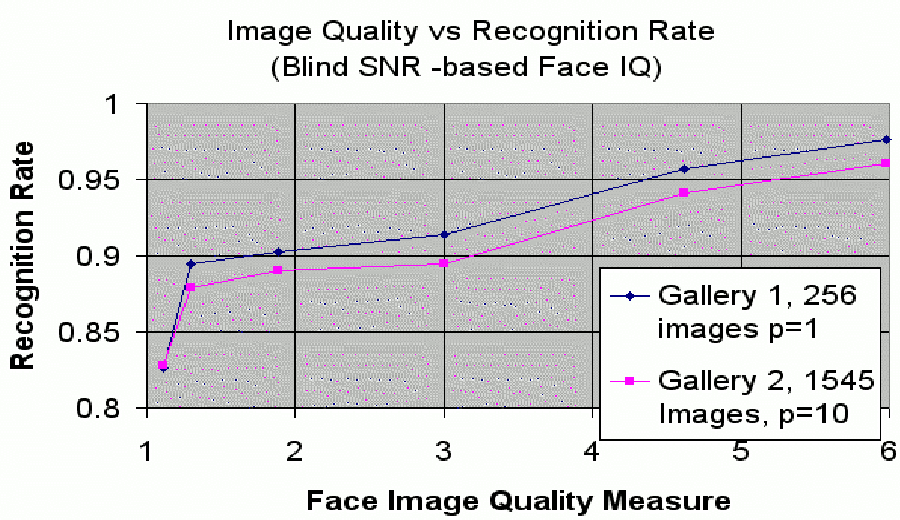
Choosing a window around eyes, define Face SNR image quality as

$$Q' = \frac{\sum \text{edge above } 2\mu\text{'s pixels}}{\sum \text{edge pixels}} \simeq \int_{2\mu}^{\infty} f_{||\nabla I||}(r)dr$$

Can also apply spatial weighting to key on eyes/nose.

Adapted from [Zhang-Blum-00].

uccs.edu



Correlations are .922 and .930 !

Also tried multiple measures of blur and contrast and multi-metric fusion. None were better than Blind SNR estimate.

Tested with FaceIt, PCA, EBGM. Generally report FaceIT

uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Why Predict Failure

- System approach – if data is not sufficient can acquire more while subject still available.
- Feedback to improve collection/sensor system.
- Decision Fusion/Boosting – can be used to weight results from multiple algorithms or multiple data sources.
- Help algorithm researchers focus on what needs "fixed"
- For "utility" qualities, task based evaluation is needed providing a "prediction", so can use it for comparison of quality measures

uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Approaches

- Input filtering – determining failure before running the classifier:
  - Using image quality to predict failure of face recognition.
- PRAT: Post Recognition Analysis Techniques
  - One example: Failure Analysis from Similarity Surface Techniques (FASST)

uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Predicting Recognition System Failure

Recognized
(as GW)

Face Recognition System

Similarity scores

Failure Prediction Engine

Recognition Successful

Recognition Failed

Rejection Successful

Rejection Failed

Rejected
(not in DB)

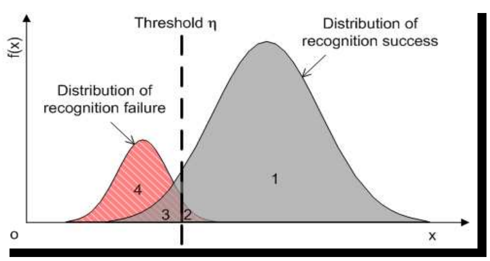
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Evaluating Failure Prediction



- Failure Prediction False Alarm Rate

$$FPFAR = \frac{|Case\ 3|}{|Case\ 3| + |Case\ 1|}$$

- Failure Prediction Miss Detection Rate

$$FPMDR = \frac{|Case\ 2|}{|Case\ 2| + |Case\ 4|}$$

	Conventional Explanation	**Prediction**	**Ground Truth**
Case 1	*True Accept*	Success	P
Case 2	*False Accept*	Success	O
Case 3	*False Reject*	Failure	O
Case 4	*True Reject*	Failure	P

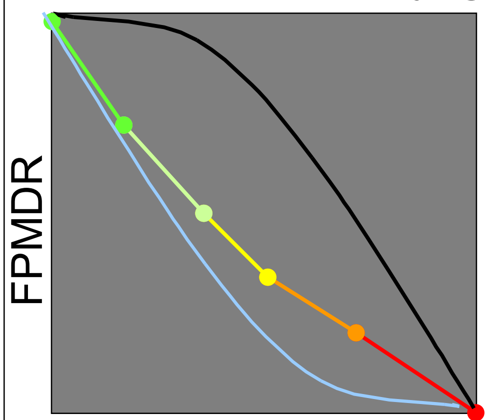
uccs.edu

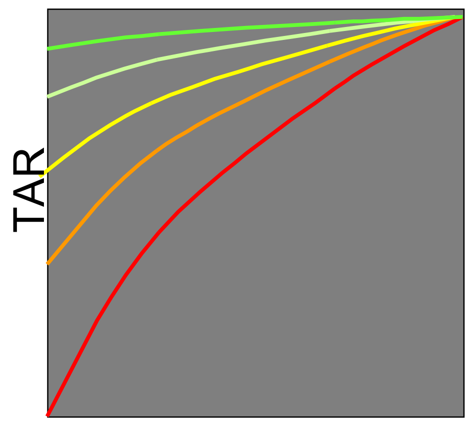
U. Colorado at Colorado Springs

UCCS

Securics

FP ROC Compared to Quality-grouped ROC





FPFAR
Uses Full Data Sets
Vary "quality" threshold

FAR
Segmenting gallery on quality inflates the difference

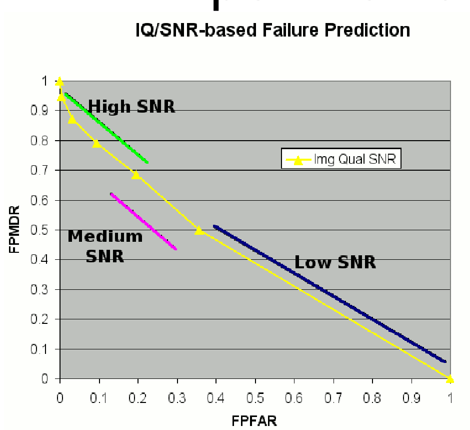
uccs.edu

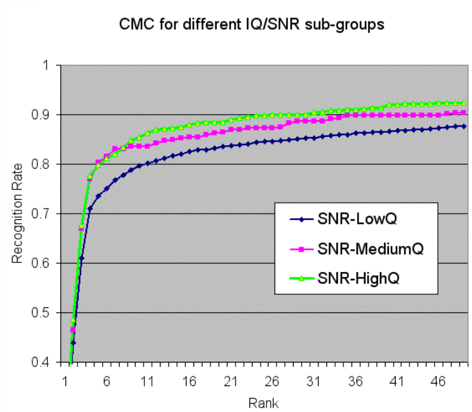
U. Colorado at Colorado Springs

UCCS

Securics

Experimental FPROC vs CMC





Constant HQ Gallery (1024), Group probes by IQ/SNR
12,000 photo-head probes

uccs.edu

U. Colorado at Colorado Springs

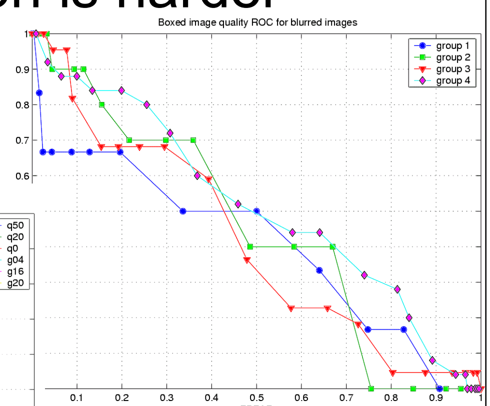
FPROC

- ✓ Allows more direct comparison of different quality measures, or a quality measure on different sensors/groups
- ± Requires an "evaluation gallery"
- ± Depends on underlying recognition system's tuning and decision making processes
- – May understate the "impact" of removing poor quality prints from process.

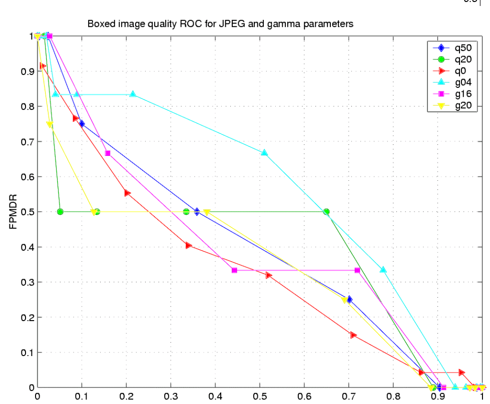
uccs.edu

U. Colorado at Colorado Springs

Quality-based Prediction is harder



Jpeg & Gamma

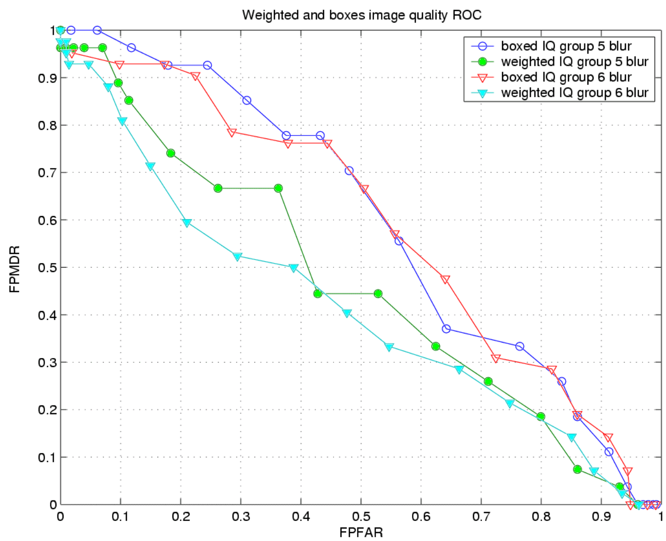


BLUR

uccs.edu

U. Colorado at
Colorado Springs

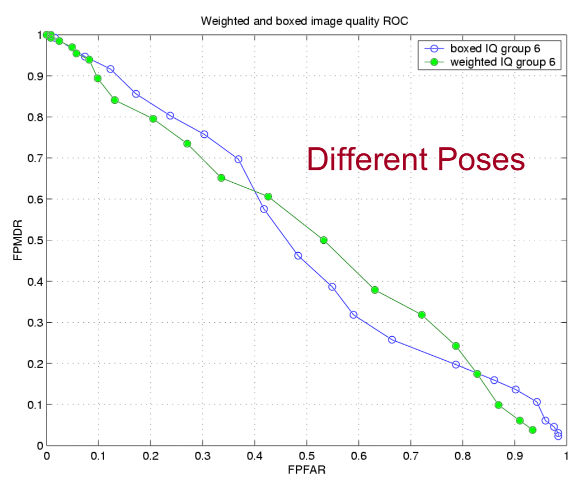
Weighting on “eyes” region helps



uccs.edu

U. Colorado at
Colorado Springs

But Probe/Gallery pose differences dominate



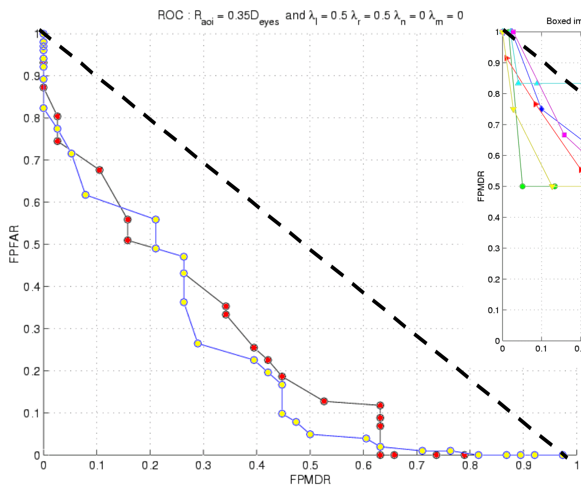
uccs.edu

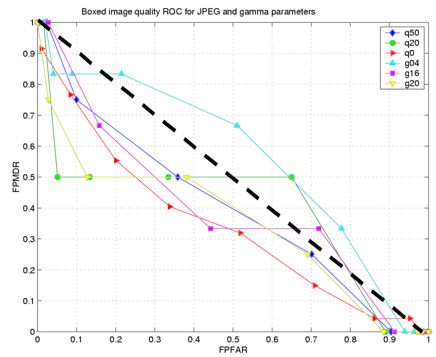
U. Colorado at Colorado Springs

UCCS

Securics

Learning added measures for Facial IQ





Compare to red lin

Add new few Features in Facial IQ based on Ada-boosted wavelets around eyes to "learn" features for eyes closes/glasses.

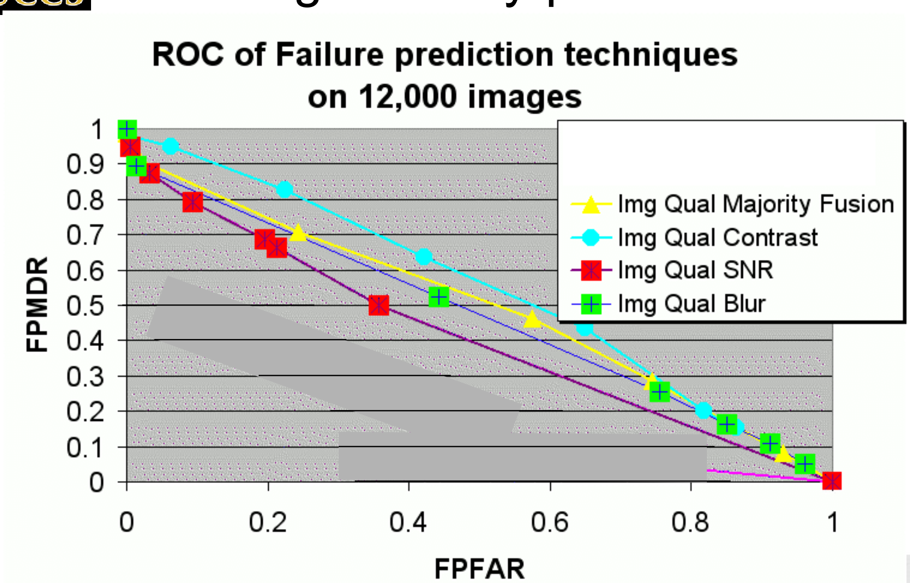
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Image Quality-predictions



uccs.edu

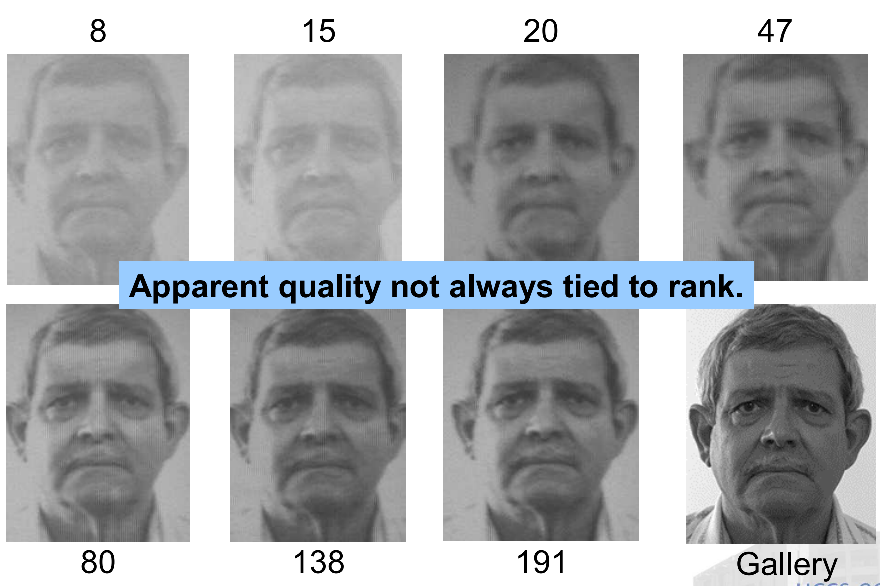
U. Colorado at Colorado Springs
UCCS

FIQ Conclusion

- Statistics of edge intensity distribution (blind image SNR estimate) are well correlated with recognition rates.
- For “good pose/lighing” images the IQ variations are fair predictor of recognition failure.
- Windowing and Weighting help as IQ becomes weak but pose and lighting are more significant.
- IQ not as good predictor when significant pose/lighting/contrast/compression variations are allowed.
- If doing “quality” should include pose/lighting estimates against “standard”

U. Colorado at Colorado Springs
UCCS

Image quality and rank



U. Colorado at Colorado Springs

100ft 200ft

Securics

rank: 6 rank: 6

U. Colorado at Colorado Springs

UCCS

Securics

PRAT: Post-Recognition Analysis Techniques

- Using data from actual recognition process, can Post Analysis predict failure?
- Many Recognition/Classification processes can be viewed using "similarity" scores.
- Failure Analysis from Similarity Surface Techniques. For details see
  - Li-Gao-Boult-05 IEEE Conf. Computational Intelligence for Homeland Security and Personal Safety, 2005
  - Riopka-Boult-05, AVBPA 2005.

uccs.edu

U. Colorado at Colorado Springs

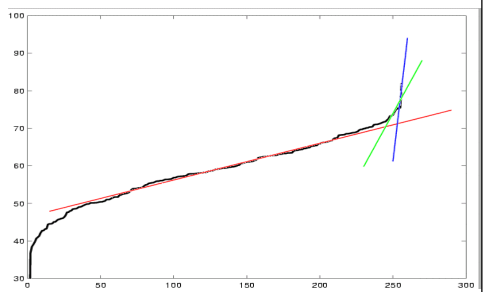
UCCS

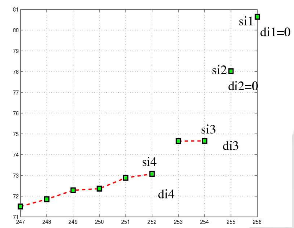
Securics

Similarity-based recognition

Failure Analysis from Similarity Surface Theory

- Similarity scores say how well target matches each DB entry.
- Used for all biometric Recognition problems
- Usually largest score is “match”. But is it good enough?
- Overall shape say a lot about if it’s a real match.





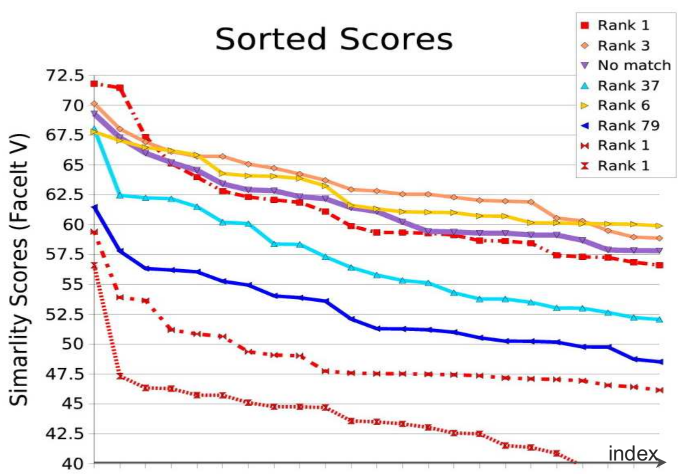
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Similarity Score Examples



Sorted similarity scores

$\{s(\mathbf{x}_i, \mathbf{y}_1), s(\mathbf{x}_i, \mathbf{y}_2), \ldots, s(\mathbf{x}_i, \mathbf{y}_n)\}$

uccs.edu

U. Colorado at Colorado Springs

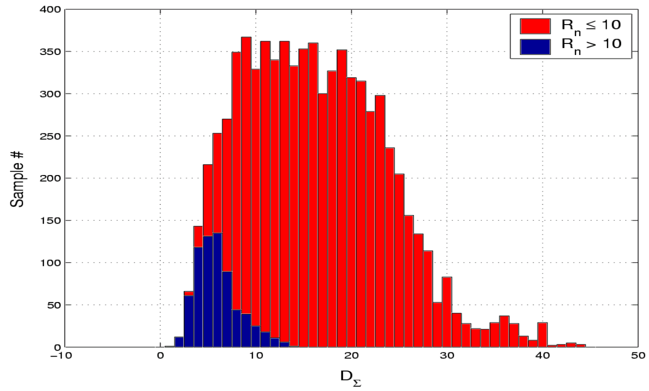
UCCS

Securics

Simple "Slope"

D_Σ = Height difference in similarity score S_1-S_p

Crude Slope estimate = D_Σ / p



- Sample size = 8,423 from *FaceIt*
- Face images from *FERET*

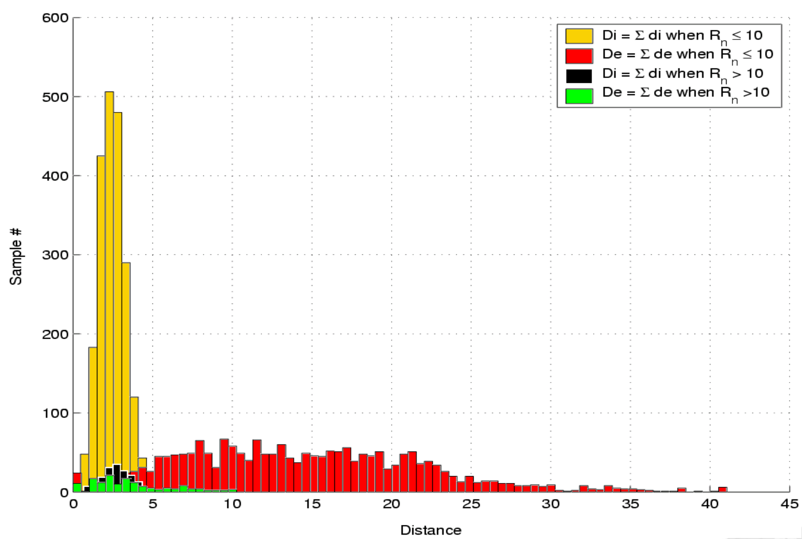
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Separation of new Measures



uccs.edu

U. Colorado at Colorado Springs

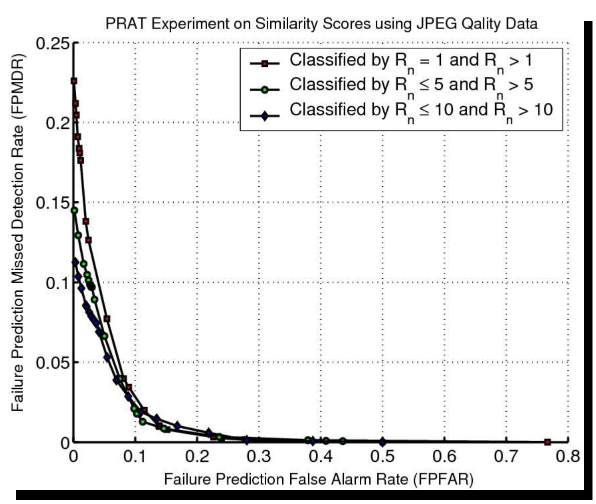
Forms of FASST tested

- Hand-chosen threshold for "slope" features (common "normalization"?)
- Ada-Boost applied to designed features of sorted similarity data of top 10% (APRAT on slides)
- 3 layer Neural Net applied to top 10% similarity + number of "gallery duplication" count

uccs.edu

U. Colorado at Colorado Springs

ROC Plots — **JPEG data**



- Sample size = 121,308 × 4
- Three partitions

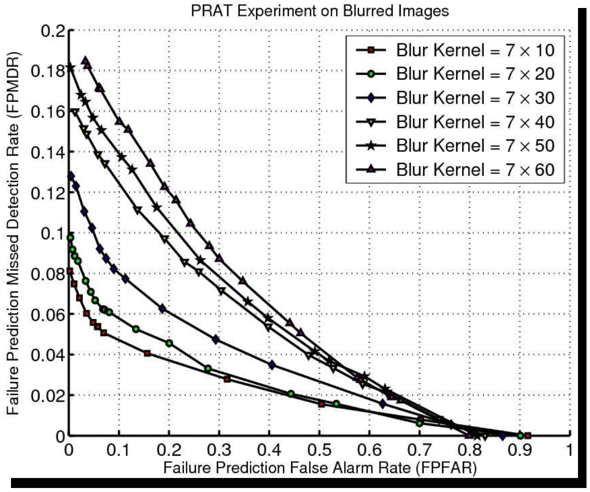
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

ROC Plots — **Blur data**



- Sample size = 4,064
- Only probe blurred

We find

- ▶ Blur kernel STD ↑ ⇒ performance ↓

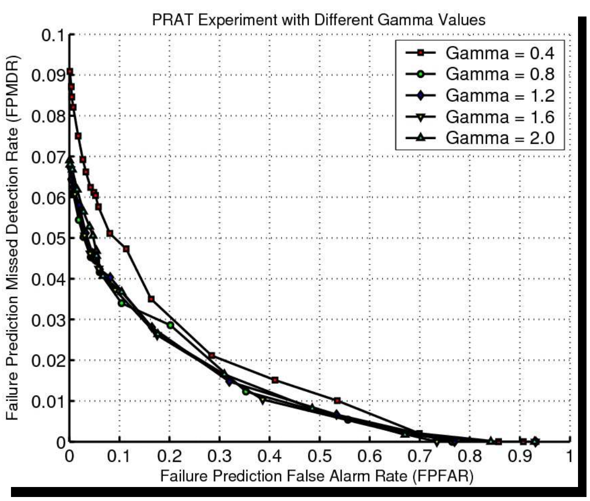
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

ROC Plots — **Gamma data**



- Sample size = 4,052

We find

- ▶ Gamma transform has little impact on prediction performance

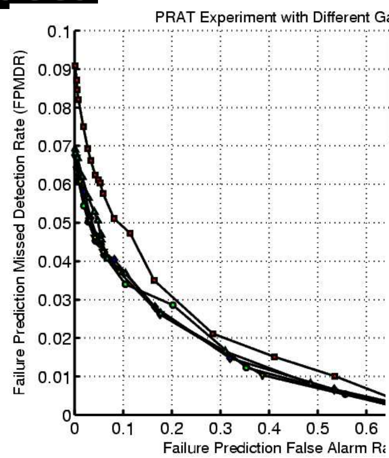
uccs.edu

U. Colorado at Colorado Springs

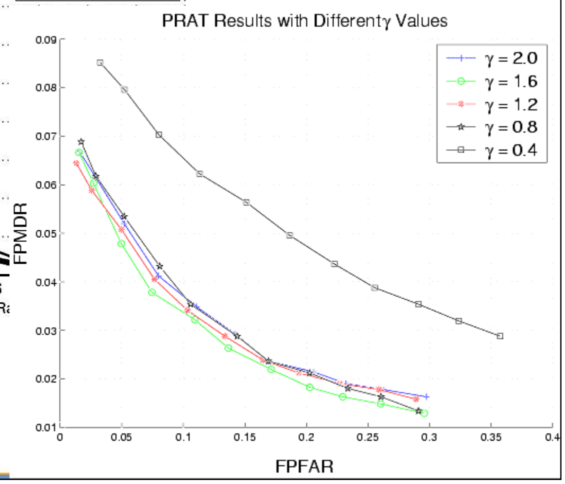
UCCS

Securics

APRAT vs PRAT (Gamma)



APRAT is good and automated!



U. Colorado at Colorado Springs

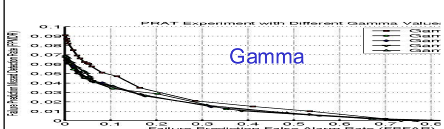
UCCS

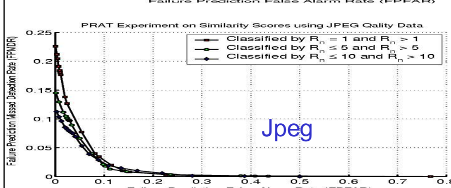
Securics

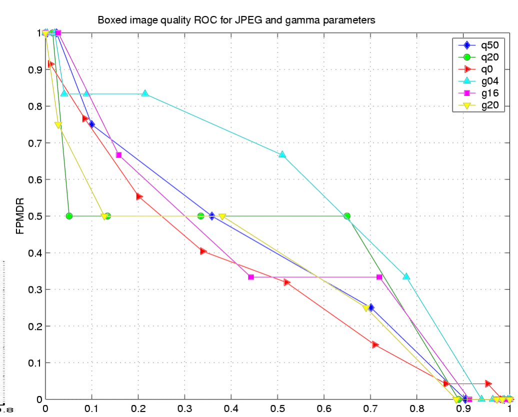
APRAT vs IQ-based prediction

APRAT

(note vertical scale!)







IQ-based

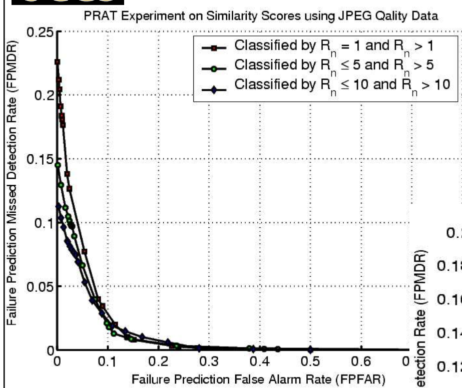
uccs.edu

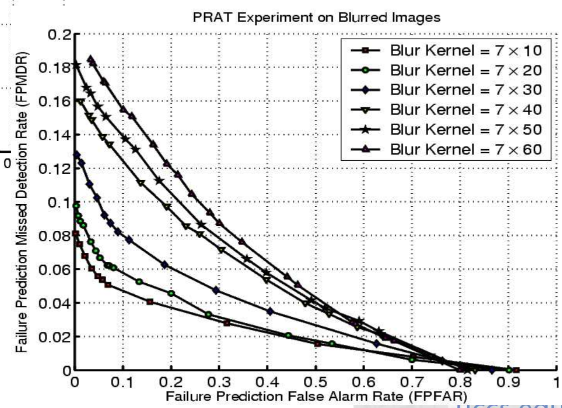
U. Colorado at Colorado Springs

UCCS

Securics

APRAT on JPEG/Blur





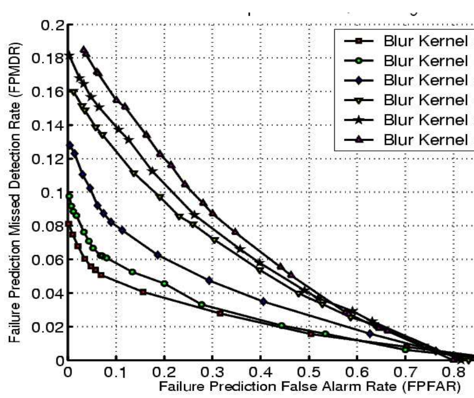
uccs.edu

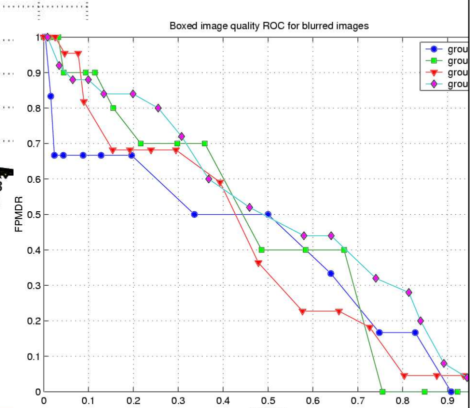
U. Colorado at Colorado Springs

UCCS

Securics

FASST vs IQ Comparison: Blur





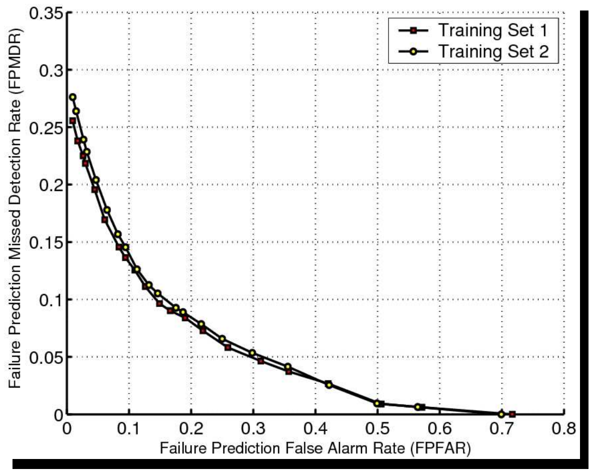
(Note vertical scale!)

U. Colorado at Colorado Springs

UCCS

Securics

ROC Plots –**Photohead data**



- Sample size = 21,353
- Cross-validation
- Real data (≈)

We find

- Predicting failure in weather more difficult
- *EER (I.e. MD=FA)* is ~12%

uccs.edu

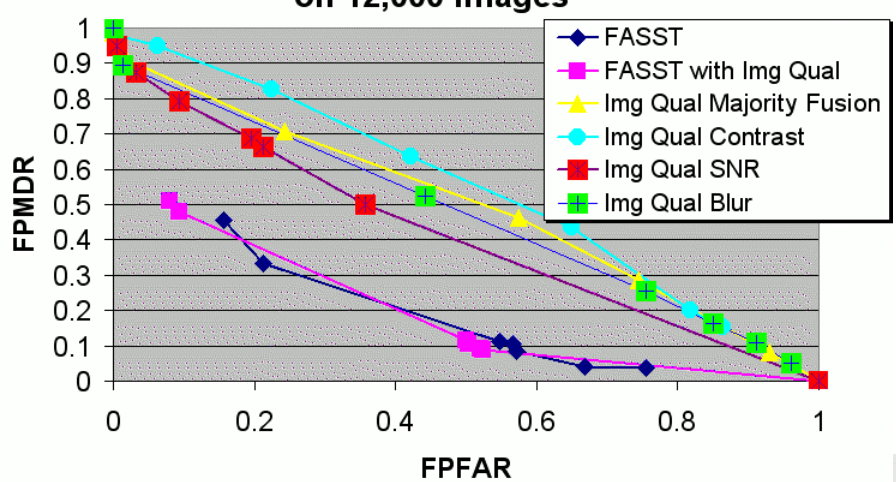
U. Colorado at Colorado Springs

UCCS

Securics

FASST and Image Quality

ROC of Failure prediction techniques on 12,000 images

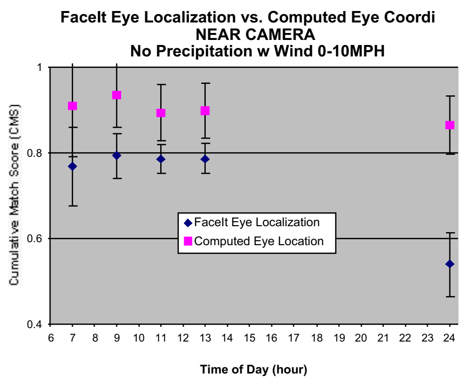


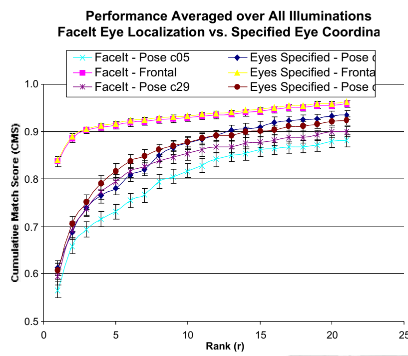
uccs.edu

U. Colorado at Colorado Springs

The Eyes Have it

- Recognition Rates unacceptable especially outdoor and at long distances.
- Riopka & Boult in ACM Biometric Workshop showed strong impact of Eye-location.



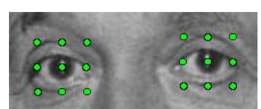


uccs.edu

U. Colorado at Colorado Springs

RandomEyes™

Predict when failure likely, and if so perturb location of features and choose best alternative.



Use a Neural Net to predict probable failure from top similarity scores.

Features for prediction:

- Eight Wavelet coefficients from a 4 point discrete Daubechies wavelet transform applied to top 8 sorted similarity scores.
- Each probe had 4 gallery images so we added two other features, number of matching IDs in top 8 and next highest rank of top ranked ID (=9 if none).

- See paper by Riopkia-Boult in AVBPA 2005

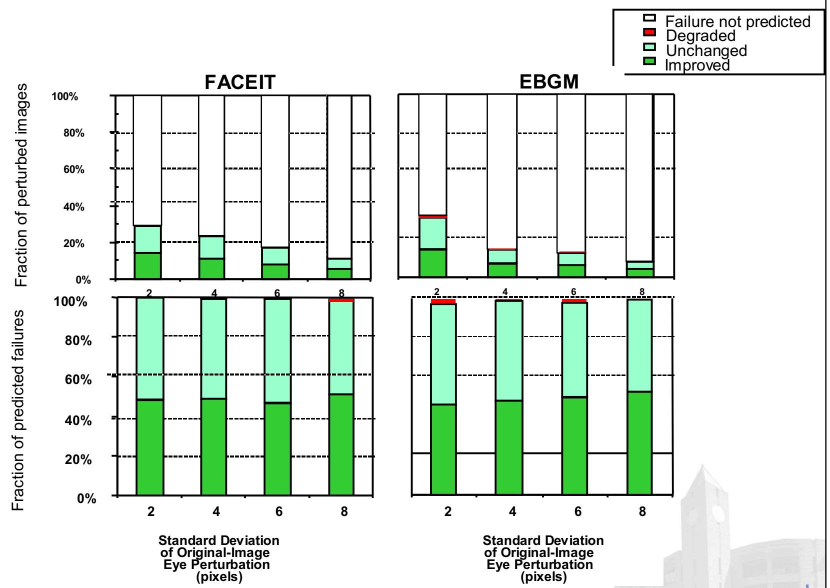
uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Synthetic Data Results



uccs.edu

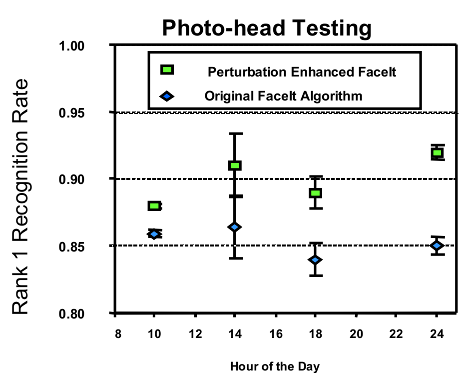
U. Colorado at Colorado Springs

UCCS

Securics

RandomEyes™ helps Photoheads

- Predicting failure and trying perturbations can significantly improve recognition



uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Conclusions/Future Work

- IQ strongly correlated to Recognition rate but a weak per image predictor. Not a good predictor when pose/lighting/eye dominates recognition rates.
- FASST, using cumulative intra-cluster distance in high ranking similarity scores is an effective predictor. Two forms on different representations/techniques show its generality.
- FASST + Image quality not significantly better
- FASST + perturbations statistically significantly improve results
- Can we apply FASST on a "test gallery" and make it useful during raw capture?
- Can FASST be useful in factor analysis and experimental assessment?

uccs.edu

U. Colorado at Colorado Springs

UCCS

Securics

Shameless plug

- Workshop on Privacy Research In Vision
- June 2005 (in conjunction with CVPR)
- Discussion oriented workshop but will have papers as well.
  - Papers due Mar 15

uccs.edu